DEVELOPMENT OF DISSIMILARITY-BASED MSPM SYSTEM

NURULSAIDATULHANIZA BINTI ZAHARI

Thesis submitted in partial fulfilment of the requirements

for the award of the degree of

Bachelor of Chemical Engineering

Faculty of Chemical & Natural Resources Engineering

UNIVERSITI MALAYSIA PAHANG

JANUARY 2014

ABSTRACT

This research is about development of dissimilarity matrix based on Multivariate Statistical Process Monitoring (MSPM) system. MSPM is an observation system to validate whether the process is happening according to its desired target. Nowadays, the chemical process industry is highly based on the non-linear relationships between measured variables. However, the conventional Principal Component Analysis (PCA) which applied based on MSPM system is less effective because it only valid for the linear relationships between measured variables. In order to solve this problem, the technique of dissimilarity matrix is used in multivariate statistical process monitoring as alternative technique which models the non-linear process which simultaneously can improve the process monitoring performance. The procedures in MSPM system consists of two main phases basically for model development and fault detection. This research focused on converting dissimilarity matrix to minor product moment before proceeding to PCA process which runs by using Matlab software. The monitoring performance in both techniques were compared and analysed to achieve the aims of this research. The findings of this study are illustrated in the form of Hotelling's T^2 and Squared Prediction Errors (SPE) monitoring statistics to be analysed. As a conclusion, the dissimilarity system is comparable to the conventional method. Thus, it can be the other alternative method in the process monitoring performance. Finally, it is recommended to use data from other chemical processing systems for more concrete justification of the new technique.

ABSTRAK

Kajian ini adalah tentang pembentukkan perbezaan matrik berasaskan sistem proses pemantauan multivariat statistik (MSPM). MSPM adalah sistem pemerhatian untuk mengesahkan sama ada proses yang berlaku mengikut sasaran yang dikehendaki. Pada masa kini, industri proses kimia adalah berdasarkan hubungan bukan linear antara pembolehubah yang diukur. Walau bagaimanapun, system konvensional Proses Analisis Komponen (PCA) yang dijalankan mengikut sistem MSPM kurang berkesan kerana ia hanya sah untuk hubungan linear antara pembolehubah yang diukur. Dalam usaha untuk menyelesaikan masalah ini , teknik perbezaan matrik digunakan dalam proses pemantauan multivariat statistik sebagai teknik alternatif yang berasaskan proses bukan linear yang pada masa yang sama boleh meningkatkan prestasi proses pemantauan. Pada dasarnya, prosedur di dalam sistem MSPM terdiri daripada dua fasa utama iaitu untuk pembentukkan model dan pengesanan masalah. Kajian ini memberi tumpuan kepada penukaran perbezaan matrik menjadi masa produk kecil sebelum bersambung ke proses PCA yang dibentuk menggunakan perisian Matlab. Prestasi pemantauan dalam kedua-dua teknik dibandingkan dan dianalisis untuk mencapai matlamat kajian ini. Hasil kajian ini digambarkan dalam bentuk pemantauan statistik Hotelling T^2 dan Squared Ramalan Kesilapan (SPE) untuk dianalisis. Kesimpulannya , sistem perbezaan matrik adalah setanding dengan kaedah konvensional . Oleh itu, ia boleh menjadi kaedah alternatif dalam melaksanakan proses pemantauan. Akhirnya, ia adalah disyorkan untuk menggunakan data dari sistem pemprosesan kimia yang lain untuk memberi justifikasi yang lebih padat berkenaan teknik baru ini..

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

X	Normal operating data
X^T	Normal operating data transpose
$\breve{X}$	Standardised data
C_{mxm}	Variance-covariance matrix
λ	Eigen values
V	Eigenvectors
P	PC scores
E	Residual matrix
N	Samples
m	Variables
i	Row
j	Column
B	Scalar product matrix
q_i	Loading vector of PCA
x	Data
$\overline{x}$	Data means
σ	Standard deviation
ϕ(x)	Nonlinear transformation
k	Principal component
A	Number of PCS retained in PCA model
n	Number of nominal process measurements per variable
$P_{i,j}$	i^{th} score for Principal Component *j*
λ_j	Eiganvalue corresponds to Principal Component *j*
z_α	Standard normal deviate corresponding to the upper (1- α) percentile
X_z	Standardized matrix of original matrix, X
I	Identity matrix
J	Centring matrix
e_i	i^{th} row in residual matrix
Q_i	SPE statistics
$\{\delta_{rs}\}$	Dissimilarity
Λ	Diagonal matrix
V^T	Normalized orthogonal matrix
α	Level of control limit

LIST OF ABBREVIATIONS

PBR	Packed bed reactor
PFR	Plug flow reactor
CA	Canonical correlation analysis
CMDS	Classical multidimensional scaling
CVA	Canonical variate analysis
FA	Factor analysis
F1	Fault 1
F2	Fault 2
F3	Fault 3
ICA	Independent component analysis
IT-net	Input-training neural network
KPCA	Kernel PCA
MDS	Multidimensional scaling
MPCA	Multi-way PCA
MSPCA	Multi-scale PCA
MSPC	Multivariate statistical process control
MSPM	Multivariate statistical process monitoring
NOC	Normal operating data
PARAFAC	Parallel factors analysis
PC	Principal component
PCA	Principal component analysis
P.D.F	Probability density function
PLS	Partial least square
SD	Singular decomposition
SVD	Singular value decomposition
SPC	Statistical process control
SPE	Squared prediction errors

CHAPTER 1

INTRODUCTION

1.1 RESEARCH BACKGROUND

The ultimate aim of any production system is to produce the maximum amount of high quality products as per requested and specified by the customers. This is regarded as highly challenging due to the nature of the processes that always change over time and are also affected by various factors such as variations of raw materials as well as operating conditions, the presence of disturbances and also modification in the process technologies. In any of the situations, one of the main critical problems is to promptly detect the occurrence of faulty or abnormal operating conditions in the routine process operation and subsequently remove them. Such issues can be addressed quite effectively by the use of process monitoring techniques. In general, there are two typical types of process monitoring schemes applied widely in chemical-based industry, which are individual-based monitoring also known as Statistical Process Control (SPC) and multivariate-based monitoring that also synonymous to Multivariate Statistical Process Control (MSPC) or Multivariate Statistical Process Monitoring (MSPM).

SPC techniques involve univariate methods, that is, observing and analysing a single variable at a time. Industrial quality problems are multivariate in nature, since they involve measurements on a number of characteristics, rather than one single characteristic. The conventional SPC charts such as Shewhart chart and CUSUM chart have been widely used for monitoring univariate processes, however they do not function well for multivariable processes with highly correlated variables. Most of the limitations of univariate SPC can be addressed through the application of Multivariate Statistical Process Control (MVSPC) techniques, which consider all the variables of interest simultaneously and can extract information on the behaviour of each variable or characteristic relative to the others. Thus, multivariate statistical process monitoring (MSPM) can be considered as the most practical method for monitoring complicated and large scale industrial processes (Manabu et al., 2000).

According to Yunus and Zhang (2010), MSPM has been shown to be a very effective process monitoring tool. The framework which has been originated from the method of statistical process control (SPC) is aimed to maintain consistent productivity by way of anticipating early warning of possible process malfunctions in the multivariate process. MSPM methods are basically algorithms that can be used for extracting important information from large multivariable data sets such as plant data. Its performance depends on how well the model describes relationships between the variables. Therefore, the key feature of such methods is the possibility to handle highly correlated, highly dimensional and noisy data. MSPM methods describe original data by the reduced set of variables which in turn makes analysis of the data much easier (Sliskovic et al., 2012).

1.2 MOTIVATION AND PROBLEM OF STATEMENT

Over last decade, many chemical process industries used MSPM as an alternative method in process monitoring performances and fault diagnosis for their plants. One of the tools in multivariable statistical techniques is Principal Component Analysis (PCA). Lindsay (2002) has defined PCA as a way to identify patterns in data and express the data in such a way to highlight their similarities and differences. PCA is a powerful tool for analysing data since patterns in data can be hard to find in data of high dimension. The other main advantage of PCA is once the patterns are found the data can be compress by reducing the number of dimensions without loss much of information.

Research done by Faezah and Athena (n.d) proved that PCA provide a roadmap to shrink a complex data set to lower dimension and it can analyse the basis of variation present in multi-dimensional data set. However, Choi, Morris and Lee (2008) said that conventional PCA based on MSPM is only valid for the non-auto correlated data with linear relationships between measured variables. Often, inefficient and unreliable process monitoring scheme can materialize as a consequence of the underlying assumption of PCA-based MSPM being violate. Recently, the chemical process industry is highly based on the non-linear relationships between measured variables. Thus, the conventional PCA based on MSPM is no longer effective for the field of the process monitoring performance and fault diagnosis in a chemical process industry.

Therefore, engineer has to find another alternative technique which can solve the current problem of the process monitoring performance and fault diagnosis in a chemical process industry to achieve good quality control expectation as the goal to produce the maximum amount of highly quality product that requested and specified by the customer. In react to this issue, dissimilarity method based on MSPM is expected to solve the current problem which models the non-linear process. Dissimilarity method is used inter distance measures which can cope either linear or non-linear process. Simultaneously, it can improve the process monitoring performance by using MSPM procedures. Thus, this research is done to study and explore about the dissimilarity and perhaps can introduce it as another alternative in process monitoring.

1.3 RESEARCH OBJECTIVES

The main aim of this research is to propose a new technique in process monitoring which applies dissimilarity-based MSPM. The dissimilarity is based on the process monitoring for non-linear multivariate processes through the application of MSPC. Therefore, the main objectives of this research are:

i. To run the conventional PCA-based MSPM system.
ii. To develop the dissimilarity-based MSPM system.
iii. To compare and analyse the monitoring performance between the conventional PCA and dissimilarity techniques.

1.4 RESEARCH QUESTIONS

i. What are the types of scales which can be used by the new system in achieving consistent process monitoring performance?
ii. How effective and efficient the new system may improve the process monitoring performance as compared to the conventional MSPM?
iii. Do the outcomes support the research aim?

1.5 RESEARCH SCOPES

The research scopes of this research are listed as follow:

i. To develop the conventional MSPM procedure in which the linear PCA algorithm is used for lowering the multivariate data dimensions.
ii. To study and explore about the dissimilarity matrix for constructing the core correlation structure.
iii. Using Matlab software platform version 7 as a tool to achieve the objectives stated earlier.
iv. Focusing on the fault detection scheme only.
v. Using Shewhart chart to monitor the process performance.
vi. Using Tennessee Eastman process as a case study.

1.6 SIGNIFICANCE OF STUDY

This study produces a new idea on how to reduce the complexity of monitoring performance by using dissimilarity matrix method in modelling all the variables involved. The method is expected to improve the monitoring progressions especially in terms of fault detection sensitiveness.

1.7 CHAPTER ORGANIZATIONS

The thesis is divided into five main chapters. The first chapter introduces the background of the research which includes the problem statement and motivation, objectives, scopes and significance of this research. The literature review is presented in chapter 2, where it describes the fundamental of MSPM, process monitoring issues and extension and multidimensional scaling in the MSPM framework. Chapter 3 explains the methodology for both conventional PCA and dissimilarity matrix methods. Chapter 4 is discussing on the result and discussion of the research and finally, conclusion and recommendations have been discussed in chapter 5.

CHAPTER 2

LITERATURE REVIEW

2.1 INTRODUCTION

Quality and safety are the two important aspects of any production process. Identification and control of chemical process is a challenging task because of their multivariate, highly correlated and non-linear nature. As mentioned in the first chapter MSPM is the effective tool in process monitoring. The aim of statistical process monitoring is to detect the occurrence and the nature of the operational change that cause the process to deviate from their main objective. This chapter is divided into five sections which are introduction, fundamental of MSPM, process monitoring issues and extension, dissimilarity in the MSPM framework and summary.

2.2 FUNDAMENTAL OF MSPM

Statistical performance monitoring of a process detects process faults or abnormal situations, hidden danger in the process followed by the diagnosis of the fault. The diagnosis of abnormal plant operation can be greatly facilitated if periods of similar plant performance can be located in the historical database (Yingwei and Yang, 2010). In general, there are four main steps of MSPM in the field of the process monitoring performance and fault diagnosis. The four main steps consist of the fault detection, fault identification, fault diagnosis and process recovery. Graphically, the steps can be viewed in an arranged manner by referring to the following flow chart in Figure 2.1:

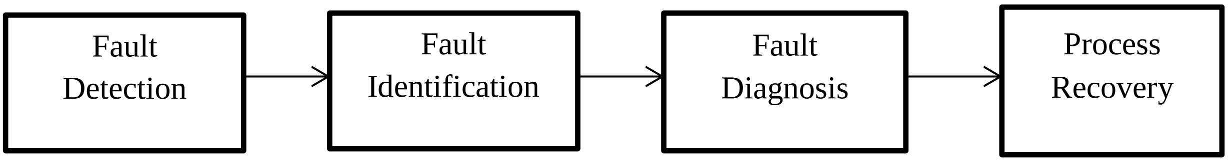

Figure 2.1: Main Steps in MSPM

Firstly, the fault detection is actually to indicate the departure of the observed sample of an acceptable range by using a set of parameters. Meanwhile for fault identification, it is to identify the observed process variables that are most relevant to the fault or malfunction which is usually identified by using the contribution of plot technique. Then, fault diagnosis is describes to determine the specific type of fault that significantly and also needs to be confirmed contributes to the signal. Finally, the process recovery is explains to remove the root of causes that contribute to the detected fault.

MSPM is based on the chemo metric techniques such as principal component analysis (PCA) and partial least squares (PLS). In previous work by Sliskovic et al. (2012), PCA was described as tool for data compression and information extraction which finds linear combination of variables that describes major trends in a data set. By using PCA, control limits are set for two kinds of statistics, T^2 and Q after a PCA model is developed. Q is the sum of squared errors, and it is a measure of the amount of variation not captured by the first few principal components. A measure of the variation within the PCA model is given by Hotelling's T^2 statistic. T^2 statistic is the sum of normalized squared scores, and it is a measure of the distance from the multivariate mean to the projection of the operating point on the subspace formed by the PCA model. PCA is also a linear transformation that is easy to be implemented for applications in which huge amount of data is to be analysed. In other words it is a numerical procedure for analyse the basis of variation present in a multi-dimensional data set (Faezah & Athena, n.d). Zhou (2010) also had described PCA is widely used in data compression and pattern matching by expressing the data in a way to highlight the similarities and differences without much loss of information. According to Spring (2010), PCA is one of techniques for taking high-dimensional data, and using the dependencies between the variables to represent it in a more tractable, lower-dimensional form, without losing too much information. The definitions of PCA from all researchers are quite similar to each other.

Based on study by Yusri (2012), first method in dimensionality reduction of PCA is a set of normal operating condition (NOC) data, X are identified off-line based on the historical process data. Then, the data are standardized to zero mean and unit variance with respect to each of the variables by using Equation (2.1) because PCA results depend on data scales.

$$\breve{\mathrm{X}}_{\mathrm{j,i}} = \frac{(\mathrm{X}_{\mathrm{j,i}} - \overline{\mathrm{X}}_{\mathrm{i}})}{\sigma_{\mathrm{i}}} \qquad (2.1)$$

Where, $\breve{\mathrm{X}}_{\mathrm{j,i}} = \text{standardized data for variiable } 'i' \text{at sample } 'j'$

$\mathrm{X}_{\mathrm{j,i}}$ = original measurement for variable 'i' at sample 'j'

$\overline{\mathrm{X}}_{\mathrm{i}} = \text{mean for variable } 'i'$

$\sigma_{\mathrm{i}} = \text{standard deviaton for variabe } 'i'$

Next, the calculation of a variance-covariance matrix, C_{mxm} by using this formula, $C = \frac{1}{n-1} X \breve{X}$ is used to develop PCA model for the NOC data. From the calculation variance-covariance matrix, the eigenvalues, λ, and eigen vectors, V can be obtained. Finally, the Principal Component (PC) scores, P can be simply develop by using this formula, $P = \breve{X}V$. The PC scores are well defined as value of the PC that has been observed for each of the n observation vectors.

2.3 PROCESS MONITORING ISSUES AND EXTENSIONS

There are various extensions have been proposed by other researchers. The process monitoring issues and extension can be divided into two categories which are process monitoring extension based on PCA and process monitoring extension based on multivariate technique which not based on PCA.

2.3.1 Process Monitoring Extension based on PCA

There are many extensions proposed by other researchers based on PCA which are Non-Linear PCA, Kernel PCA, Multi-Way PCA, Multi-Scale PCA and others. In this research, only three process monitoring extensions based on PCA will be described in more details, which include Non-Linear PCA, Multi-Scale PCA and Kernel PCA.

Nikolov (2010) proposed that Non-Linear PCA is one of the process monitoring extensions based on linear technique of PCA. There several approaches to dealing with nonlinear datasets within the framework of PCA. One possibility is to model the data with a mixture of principal component analysers that trace out the nonlinear distribution using multiple linear principal subspaces. Assuming a Gaussian distribution for each subspace, the probability of a given data point is then defined by the probability each subspace assigns to the point and the probabilities that the point belongs to each subspace.

In Non-linear PCA, the Input-Training network has been developed to reduce the network complexity (Tan & Mavrovouniotis, 1995). There are three basis steps to form the work. Firstly, the Linear PCA is used to perform the linear transformation in which the observation is rotated to a new set of uncorrelated ordinates permitting the main linear information to be extracted and condensed at the same time while maintaining sufficient data variance in the transformed data, so that the non-linear correlations is not excluded from the model. Next, the linear PC scores are rescaled to unit variance to enable the recovery of the non-linear structure in the new ordinates space of the transformed data. Finally, network optimization is improved through the use of Levenberg-Marquardt algorithm to interpret the non-linear structure in the transformed data.

Other extensions of PCA are Multi-Scale PCA (MSPCA) which is the nature of MSPCA makes it appropriate to work with the data is usually not fixed and represent the cumulative impact of many underlying process phenomena which each operating at different scale. The MSPCA methodology consists of decomposing each variable on a selected family of wavelets. The PCA model is then determined independently for the coefficients at each scale. The models at important scales are then combined in an efficient scale-recursive manner to yield the model for all scales together. For multivariate statistical process monitoring by MSPCA, the region of normal operation is determined at each scale from data representing normal operation. For new data, the important scales are determined as those where the current coefficient violates the detection limits. The actual state of the process is confirmed by checking whether the signal reconstructed from the selected coefficients violates the detection limits of the PCA model for the significant scales (Bakshi, 1998). Study done by Vijaykumar et al. (2012) shown that the multi-scale principal component generalizes the usual PCA of a multivariate signal seen as a matrix by performing simultaneously a PCA on the matrices of details of different levels. In addition, a PCA is performed also on the coarser approximation coefficients matrix in the wavelet domain as well as on the final reconstructed matrix. By selecting conveniently the numbers of retained principal components, interesting simplified signals can be reconstructed.

Besides that, Kernel PCA (KPCA) has been proposed by Kruger, Zhang & Zie (n.d) as one of PCA extensions. In construct the kernel matrix, a nonlinear transformation $\phi(x)$ from the original D-dimensional feature space to an M-dimensional feature space, where usually $M > D$. Then each data point x_n is projected to a point $\phi(x_n)$. Traditional PCA can be performs in the new feature space, but this might be extremely costly. Thus kernel methods are used to simplify the computation (Wang, 2012). The main benefit is that the original nonlinear behaviour can be mapped into the feature space and then analysed through linear correlation (through a specified means of kernel function), and as a result, linear PCA can be effectively executed for monitoring.

2.3.2 Process Monitoring Extension based on Multivariate Technique

In this literature review will explain more detail only three process extension based on multivariate technique. There are Partial Least Square (PLS), Independent Component Analysis (ICA) and Canonical Variate Analysis (CVA). Actually, there are many types of extensions based on multivariate technique includes Parallel Factors Analysis (PARAFAC), Canonical Correlation Analysis (CA) and Factor Analysis (FA) which not discusses in this literature.

Yusri (2012) stated that Partial least square (PLS) is the main competitor of PCA with regard to its popularity in the area of MSPM application. Among others, the original works have been proposed by Nomikos and MacGregor, (1995), as well as Kourti et al., (1995), for batch process monitoring using multi-way PLS, whereas Kourti and MacGregor, (1995) proposed using PLS for both continuous and batch processes. PLS regression is a recent technique that generalizes and combines features from principal component analysis and multiple regressions. It is particularly useful to predict a set of dependent variables from a very large set of independent variables. The goal of PLS regression is to predict Y from X and to describe their common structure. When Y is a vector and X is full rank, this goal could be accomplished using ordinary multiple regression. When the number of predictors is large compared to the number of observations, X is likely to be singular and the regression approach is no longer feasible (Abdi, n.d). In such cases, although there are many factors, there may be only a few underlying or latent factors that account for most of the variation in the response. The general idea of PLS is to try to extract these latent factors, accounting for as much of the manifest factor variation as possible while modelling the responses well.

Generally, Independent Component Analysis (ICA) is statistical technique for expose the secret factor that underlying a set of random variables, measurements or signals. ICA identifies non-Gaussian components which are modelled as a linear combination of the biological features. These components are statistically independent such as there is no overlapping information between the components. ICA therefore involves high order statistics, while PCA constrains the components to be mutually orthogonal, which involves second order statistics. As a result, PCA and ICA often choose different subspaces where the data are projected. As ICA is a blind source signal separation, it is used to reduce the effects of noise or artefacts of the signal since usually noise is generated from independent sources (Yao**,** Coquery and Kim, 2012)**.** According to the study by Matei (n.d), there are two distinct approaches towards computing the ICA. One employs high order cumulant and is found mainly in the statistical signal processing literature and the other uses the gradient-descent of non-linear activation functions in neuron-like devices and is mainly developed in the neural networks community. Each of the above approaches has advantages and shortcomings: the computation of high order cumulants is very sensitive to outliers and lack of sufficient support in the data especially for signals having a long-tailed probability density function (p.d.f.), while the neural-networks algorithms may become unstable, converge slowly and most often require some extra knowledge about the p.d.f. of the source signals in order to choose the non-linearities in the neurons.

Another extension of process monitoring based on multivariate technique is Canonical Variate Analysis (CVA). According to Simoglou, Martin and Morris (2002), the concept of PLS is quite similar to CVA which is in the method of linear combine calculation of past values of the system input or output that are most highly correlated with linear combine of the future of the outputs process. CVA give an advantage compared to other technique which is in terms of model stability and parsimony for example, CVA only required fewer identified parameter in the final models. CVA can provide more rapid detection when comparing CVA with PLS based on process monitoring schemes.

2.4 DISSIMILARITY IN THE MSPM FRAMEWORK

In the present work, in order to improve the performance of process monitoring, a new statistical process monitoring method is proposed. The proposed method is based on the idea that a change of operating condition can be detected by monitoring a distribution of time-series data, which reflects the corresponding operating condition. In order to quantitatively evaluate the difference between two data sets, a new index representing dissimilarity is defined. According to Manabu et al. (2000), concept of dissimilarity is used for classifying a set of data for example, the degree of dissimilarity between two classes is measured by the distance between barycentre of the data and two classes with the smallest degree of dissimilarity are combined for generating a new class.

Based on the study of Yunus and Zhang (2010), classical multidimensional scaling (CMDS) is another technique which used compressing multivariate data by using dissimilarity measures for process monitoring. This technique actually is same used in this research. In this work, the dissimilarity measures have been particularly constructed based on two different scales, city block and mahalanobis distances, which are shown respectively by equation (2.2) and (2.3) (Cox et. al., 1994):

City block distance: $$\delta_{rs} = \sum_i |x_{ri} - x_{si}| \tag{2.2}$$

Mahalanobis distance: $$\delta_{rs} = \{(x_r - x_s)^T \textstyle\sum_i^{-1}(x_r - x_s)\}^{1/2} \tag{2.3}$$

The algorithm for finding the dissimilarity can be summarized as (Borg and Groenen, 2005):

$$\mathbf{A} = [\delta_{rs}^2] \tag{2.4}$$

$$\mathbf{B} = -\frac{1}{2}\ \mathbf{J\ A\ J} \tag{2.5}$$

$$\mathbf{B} = \mathbf{V\Lambda V^T} \tag{2.6}$$

Matrix **A** contains the squared dissimilarities. Then **A** is doubly centred using the centring matrix $\mathbf{J} = \mathbf{I} - \frac{\mathbf{11'}}{n}$ and multiplied by -1/2 to form matrix **B**. Then **B** is expressed in terms of its spectral decomposition, $\mathbf{V\Lambda V^T}$, where $\mathbf{\Lambda}$ is the diagonal matrix of ordered eigenvalues of **B**, **V** the matrix of corresponding eigenvectors.

Moreover, a search was also carry out for investigating the correlation between PCA and dissimilarity. This relationship is viewed from the close fundamental algorithms between conventional PCA and dissimilarity procedures. Cox et. al. (1994) had described the relationship between minor product moment and dissimilarity matrix by using algorithm manipulations approach. They started the procedure by defining the scalar product matrix, **B**, $\mathbf{B} = \mathbf{XX^T}$, in which **X** is standardized NOC data. By applying the Singular Decomposition (SD) operation on **B**, the following are obtained:

$$\mathbf{Bu_i} = \boldsymbol{\lambda}_i\mathbf{u_i} \tag{2.7}$$

$$\mathbf{XX^Tu_i} = \lambda_i\mathbf{u_i} \tag{2.8}$$

Multiplying both side with $\mathbf{X^T}$

$$\mathbf{X^T\,[XX^Tu_i] = X^T\,[\lambda_i u_i]} \tag{2.9}$$

By which,

$\mathbf{C} = \mathbf{X}^{\mathbf{T}}\mathbf{X}$; $\mathbf{C}$ represent the minor product moment

$\mathbf{q_i} = \mathbf{X}^{\mathbf{T}}\, \mathbf{u_i}$; $\mathbf{q_i}$ represent loading vector of PCA

So,

$$\mathbf{C q_i} = \boldsymbol{\lambda_i q_i} \tag{2.10}$$

By embedding the algorithm of the conventional PCA through dissimilarity, it may provide variety of results in terms of configuration plots for process monitoring. This is because the result can figure out both linear and non-linear relationships measured variables.

2.5 SUMMARY

As a conclusion, there are four main steps in MSPM in the field of the process monitoring performance and fault diagnosis which are fault detection, fault identification, fault diagnosis and process recovery. This research focuses more to the fault detection. The conventional PCA is the one of the basic technique in MSPM. The definition of PCA is a statistical method for dimensionality reduction of the quality variable space. Besides that, there two types of process monitoring issues and extension which are process monitoring extension based on PCA and process monitoring extension based on multivariate technique. Extension based on PCA includes Non-Linear PCA, Multi-Scale PCA and Kernel PCA, while, extension based on multivariate technique are Partial Least Square (PLS), Independent Component Analysis (ICA) and Canonical Variate Analysis (CVA). It may provide variety of results in terms of configuration plots for process monitoring by embedding the algorithm of the conventional PCA through dissimilarity.

CHAPTER 3

METHODOLOGY

3.1 INTRODUCTION

This chapter will illustrate procedures on MSPM through development of PCA and dissimilarity matrix methods. Generally, there are varieties of technique in multidimensional scaling (MDS). It includes classical scaling, non-metric scaling, procrustes analysis, biplot and general dissimilarity. This chapter can be divided into three sections which are introduction, methodology and summary.

3.2 METHODOLOGY ON DISSIMILARITY-BASED MSPM

In this research, the main focuses of the methodology is fault detection in MSPM system. According to Mason and Young (2002), the complete procedures of fault detection consists of two main phases namely as off-line modelling and monitoring (Phase I) and on-line monitoring (Phase II):